Fanny Sarrazin, PhD
Université Paris-Saclay, INRAE, HYCAR Research Unit
1 rue Pierre-Gilles de Gennes
92160 Antony, France
fanny.sarrazin@inrae.fr

Antony, 20^{th} July 2024,

Dear Dr Popp,

Please find enclosed a revised version of our manuscript "Gridded dataset of nitrogen and phosphorus point sources from wastewater in Germany (1950–2019)" in which we address the reviewer's comments. We wish to thank you, Maryna Strokal and Olga Vigiak for appreciation of our manuscript. We are grateful to the two reviewers for their suggestions that helped us to further improve the manuscript. Below in this document, we provide a point-by-point reply to the reviewers, and we summarize here the main changes to the manuscript:

- We expanded our analysis of parameter uncertainty in Section 5.1 and in particular we added a table that shows the reduction in parameter ranges (Table 5) in response to the comments of Olga Vigiak.
- We substantially edited the manuscript to make it more concise and improve readability (in particular in the data section 3) in response to the comments of Olga Vigiak.
- We clarified a number of points in the manuscript, in particular regarding the objective of our analysis at river basin scale and rural emissions, in response to the comments of Maryna Strokal.
- We corrected a mistake in Equations 2 and 6.

Thank you for your consideration.

Yours sincerely,

Fanny Sarrazin

Review by Maryna Strokal

Overall comment: *The authors did a good job in developing and providing the gridded datasets for point source emissions of N and P. the manuscript is well written with good structure. Complements to the authors on balancing well between details and general descriptions. It is easy to follow the methodology. I appreciate that the authors analyzed the uncertainties in such detailed datasets. I have a few points that could improve the relevance and importance of this work:*

Reply: We thank the reviewer for their appreciation of our manuscript and for the useful comments and suggestions for improvement. In our replies below, line numbers refer to the revised version of our manuscript (version without track-changes).

Comment 1: *1. The data sources were mainly at the NUTS-1 scale. The emissions of point sources of N and P are downscaled to grids. Uncertainties are analyzed at the river basin scale. It is an interesting choice for the scales. It would be good to reflect on this choice, especially on the choice for the gridded emissions, but uncertainties in those emissions are analyzed at the river basin. Why was that choice made? How can this basin scale uncertainty analysis build trust in modeled gridded emissions considering the data sources at the NUTS-1 level?*

Reply 1: Thank you for this comment. In Section 5.3, we analyse the uncertainties at the grid level and we acknowledge that they may be large. It is however useful to provide the data at grid level to give the flexibility to users to aggregate the point sources estimates at any spatial scale of interest. In this sense, we believe that it is interesting to examine the uncertainty at different spatial aggregation levels to guide future uses of the developed dataset. We selected river basins as the scale of aggregation because it is a scale of interest for water quality studies. To make this point clearer, in our revised manuscript we made the following changes:

- at the very end of the section 1 (L. 113–115), we now provide a better explanation of the objective of the uncertainty analyses with the following sentence: " We discuss the uncertainties of our point source estimates at grid and river basin level to guide future uses of the dataset for water quality studies.".

- in section 5.4, we modified the first sentence L. 726–727 to better link with the uncertainty analysis at grid level in section 5.3 (we report in the following the addition in bold): "**Given the substantial uncertainties in the spatial pattern of the point sources at high spatial resolution (grid level, Sect 5.3)**, we examine the uncertainties in the point source estimates at river basin level, which is of interest to water quality studies, as further discussed in Sect. 6.1. "

- in section 6.1 L829–832 we had already indicated the following sentences in the first version of the manuscript: "Moreover, the uncertainties in our estimates decrease from grid level (Fig. 8) towards larger spatial aggregation units (e.g. river basins in Fig. 9). In particular, for earlier years, using the data at larger spatial scales of aggregation (above 100 km^2) is more reliable given the differences between the two downscaling schemes."
  In the revised version we added after that (L. 832–834) the following text: "Therefore,

using the dataset directly at grid level (Sect. 5.3) may be prone to large uncertainties and the analysis at river basin level (Sect. 5.4) allows reflection on suitable spatial aggregation scales for water quality studies. In this respect, our gridded dataset provides the flexibility to aggregate the point sources estimates at any spatial level of interest.".

Comment 2: *2. The role of rural point source emissions is important. Germany is a country with a lot of sewage systems in urban and rural areas. It would make the paper stronger if more discussion is provided on how rural emissions are considered, and the role of rural sanitation in those emissions. This might be interesting for other countries. Some countries do not have a lot of rural sewage systems. In this case: how can the proposed methods be still useful?*

Reply 2: Thank you for this comment. We agree that the emissions from the population not connected to the sewer system and/or wastewater treatment plants (WWTPs) can be an important source of N and P contamination. Our study aims at quantifying grid level urban and rural emissions that contribute to point sources, including both treated and untreated emissions as stated in Section 1 L111-112: "Our dataset encompasses emissions treated in urban WWTPs, including domestic and industial (indirect) emissions, as well as untreated domestic emissions collected in the sewer system". We also consider emissions from the population not connected to the sewer system, but whose wastewater is collected in cesspits (sealed tanks) and transported by trucks to WWTPs (see L198–199).
Furthermore, we estimated the gross emissions at NUTS-1 level for the remaining of the population which is not connected to sewer or WWTPs (this covers not only the rural population but also the urban population for earlier years, as can be seen from Supplementary Fig. S21-S24). Our dataset includes these NUTS-1 level information (`https://doi.org/10.5281/zenodo.10500535`). We see from Fig. 7 that these emissions are substantial in the past (in the 1950s), and that their importance then decline with time (as we explain in Sect. 5.2). In this respect, we recognize the importance of the emissions from disconnected population even in a country with an advanced wastewater handling system like Germany.
However, the fate of these gross emissions from disconnected population is uncertain and they can be either a diffuse or a point source. Due to a lack of detailed information on these emissions, previous studies made simplifying assumptions to account for these emissions in Germany [Fuchs et al., 2010], over Europe [Grizzetti et al., 2022, Vigiak et al., 2020, 2023] and globally [Morée et al., 2013, Van Puijenbroek et al., 2019]. Unravelling the fate of the emissions from disconnected population is beyond the scope of this study. In this regard, in Section 6.2 of the first version of our manuscript, we recognize that future studies should strive to improve the estimation of the emissions from disconnected population.
In our revised manuscript, we expand this discussion, as follows (L 866–874): "Another potentially important contributor to point sources is the domestic emissions that are not connected to the sewer system nor to WWTPs via transport from cesspits. While these emissions are overall of limited importance in the recent period, their magnitude is large in the earlier period (Sect. 5.2). It would be therefore critical to elucidate their fate in Germany and in other countries where these emissions can be substantial for the recent period as well [Vigiak et al., 2020]. These emissions are handled in particular in septic tanks or independent wastewater systems [Vigiak et al., 2020, 2023]. They could be either a diffuse source to soils or a point source to

surface waters, as documented for example in MUGV (2010) for the recent period in Germany. Due to a lack of detailed information on these emissions, previous studies made simplifying assumptions. In Germany, Fuchs et al. [2010] consider that disconnected population is equipped with septic tanks from which a part of the N and P is transported to WWTPs, while the other part is a diffuse source. In Europe, Grizzetti et al. [2022] and Vigiak et al. [2020] consider that it is entirely a diffuse source and that N and P are reduced with the same efficiency as that of primary treatment [Vigiak et al., 2020] or possibly secondary treatment [Grizzetti et al., 2022]. Globally, Morée et al. [2013] and Van Puijenbroek et al. [2019] consider that the urine part is a point source, while the feces part is a diffuse source."

As we discussed in Section 6.2, other N and P emission pathways would also require further investigation in future studies. We propose the use of sensitivity analysis as a way of assessing the impact of different assumptions in water quality assessments. To further reflect on this aspect, we added the following text at the end of Section 6.2 in the revised manuscript (L.896–898): "Overall, we propose that future water quality studies could perform sensitivity analysis to better understand the impact of different possible assumptions on the N and P pathways discussed in this section. For Germany, such investigation is facilitated as we provide all data that we produced, along with our model code (see code and data availability section)".

Comment 3: *3. This point is a bit also related to the previous: the manuscript would benefit from a discussion the applicability of the proposed methods for other regions and countries. countries differ in their urban and rural waste management. On top of this, not all countries have such detailed datasets at the NUTS-1 levels. This is especially true for developing countries. Can the proposed methods be used for those developing countries, if yes, what needs to be adjusted? if not, why? what would be alternatives?*

Reply 3: We build on a methodology for point sources estimation that was used at a large scale, namely over Europe [Vigiak et al., 2020] and globally [IPCC, 2019, Morée et al., 2013, Van Drecht et al., 2009]. We take the opportunity of having detailed data for Germany (sub-national statistics and observational data of wastewater treatment plant emissions) to improve these previous large-scale point sources estimations. This is stated in Section 1 at L107–109: "We use a modelling approach that builds in particular on Morée et al. [2013], Van Drecht et al. [2009], Vigiak et al. [2020] and IPCC [2019], while we make use of observational data of WWTP N and P emissions to constrain our modeled estimates and check their plausibility.".

In this respect, we modified the last sentence of the revised manuscript where we call for the collection and processing of further data where available to improve point sources estimation, similar to our study. The text now reads as follows (L923–926): "A similar approach could be adopted by other researchers to develop other national and regional datasets where sub-national and observational point sources datasets are available. This would contribute to improve large-scale understanding of nutrient point sources and their impact on the (aquatic) environment."

Comment 4: *4. Some detailed comments: - Please clarify the forms of N and p that are modeled. Please also justify the choice for those forms.*

Reply 4: Thank you for this remark. We consider total N and P and do not model specific N forms, similar to previous studies [Morée et al., 2013, Van Drecht et al., 2009, UBA, 2020, Vigiak et al., 2020, 2023]. Although separating the different N and P species would be highly valuable, it would also require substantial additional work, which is beyond the scope of this study. In Section 6.2 (L888–895), we discuss the limits of this simplification and gave some first insights on how this could be addressed in future studies. To make clearer the fact that we consider total N and P in our revised manuscript, we added the following sentence in this introduction (L. 109–111): "As in previous studies [Morée et al., 2013, Van Drecht et al., 2009, UBA, 2020, Vigiak et al., 2020, 2023], we assess total N and P without distinction between the different forms of N and P."

Comment 5: *- Table 1 has lower and upper bound. it would also good to add mean or median*

Reply 5: We actually sampled the parameters from a uniform distribution, as stated in Section 4. This implies that the mean/median values can be simply retrieved as the mean of the lower and upper values reported in Table 1. Therefore, we think it would be redundant to add the mean/median values in the table. In our revised manuscript, we added a note in the caption of Table 1 to make this clearer: "The parameters are sampled from a uniform distribution".

Comment 6: *- Why point sources? Diffuse sources are as important as point sources and are more difficult to control.*

Reply 6: As we state in the introduction section (L57–58), both N and P point and diffuse sources are important as far as water quality is concerned. Since we cannot treat the two sources in a single paper, we chose here to focus on point sources. Note that, in previous studies, we made available data of N diffuse sources (N surplus) for Germany at river basin level [Ebeling et al., 2022] and over Europe at grid level [Batool et al., 2022]. In our revised manuscript, we now explicitly mention in Section 6.1 that our dataset complements these existing diffuse sources datasets (L. 805–807).

Comment 7: *- Why Germany? Can we learn from this exercise for other countries?*

Reply 7: We explain in the introduction section about the importance of developing a long-term consistent point sources dataset for Germany (L97–100): "This is crucial to inform water quality strategies in Germany where the majority of the national monitoring sites for flowing surface water have shown nitrate and phosphorus concentrations above a limit that would ensure a good ecological status (for instance 81 % for nitrate and 70 % for phosphorus in 2015, Arle et al., 2017). Furthermore, N and P emissions in Germany have contributed to the eutrophication of the North and Baltic Sea since the mid-twentieth century (EEA et al., 2019; Arle et al., 2017)." With our study, we demonstrate how detailed data can be used to provide improved point sources estimates. We refer to our reply to Comment 3 above, where we explain that we take the opportunity of having detailed data for Germany (sub-national statistics and observational data of wastewater treatment plant emissions) to improve previous large-scale point sources estimations. We call for the collection and processing of further data where available to improve point sources estimation, similar to our study.

References

Stephan Fuchs, Ulrike Scherer, Ramona Wander, Horst Behrendt, Markus Venohr, Dieter Opitz, Thomas Hillenbrand, Frank Marscheider-Weidemann, and Thomas Götz. *Calculation of Emissions into Rivers in Germany using the MONERIS Model.* Texte—46/2010. German Environmental Agency (UBA), Berlin, Germany, 2010. URL https://www.umweltbundesamt.de/en/publikationen/calculation-of-emissions-into-rivers-in-germany. last access: 26 January 2023.

B Grizzetti, O Vigiak, A Udias, B Bisselink, A Pistocchi, F Bouraoui, A Malagó, A Aloe, M Zanni, F Weiss, J Hristov, J Wilson, E Pisoni, A De Meij, A De Roo, M Macias, and A Stips. *A European assessment of freshwater availability and nutrient pollution : historical analysis and scenarios developed in the project Blue2.2.* Joint Research Centre JRC130025. Publications Office of the European Union, Luxembourg, 2022. doi: 10.2760/924432.

Olga Vigiak, Bruna Grizzetti, Michela Zanni, Alberto Aloe, Chiara Dorati, Faycal Bouraoui, and Alberto Pistocchi. Domestic waste emissions to European waters in the 2010s. *Scientific Data*, 7:33, 2020. doi: 10.1038/s41597-020-0367-0.

Olga Vigiak, Bruna Grizzetti; Alberto Aloe, Michela Zanni, Faycal Bouraoui, and Alberto Pistocchi. *Domestic waste emissions to European freshwaters 1990 - 2016 (v. 1.0).* European Commission, Joint Research Centre [data set], 2023. URL http://data.europa.eu/89h/104da345-4ba1-444b-8df4-97bb57950a8e. last access: 02 October 2023.

A. L. Morée, A. H. W. Beusen, A. F. Bouwman, and W. J. Willems. Exploring global nitrogen and phosphorus flows in urban wastes during the twentieth century. *Global Biogeochemical Cycles*, 27(3):836–846, 2013. doi: 10.1002/gbc.20072.

P.J.T.M. Van Puijenbroek, A.H.W. Beusen, and A.F. Bouwman. Global nitrogen and phosphorus in urban waste water based on the shared socio-economic pathways. *Journal of Environmental Management*, 231:446–456, 2019. doi: 10.1016/j.jenvman.2018.10.048.

MUGV (Ministerium für Umwelt, Gesundheit und Verbraucherschutz Brandenburg). *Wegweiser für den Einsatz von Kleinkläranlagen und Sammelgruben.* Potsdam, 2010. URL https://mluk.brandenburg.de/sixcms/media.php/9/Wegweiser-Kleinklaeranlage-Sammelgruben.pdf. last access: 19 May 2024.

IPCC. Chapter 6 - Wastewater Treatment and Discharge. In E. Buendia, K. Tanabe, A. Kranjc, J. Baasansuren, M. Fukuda, Ngarize S., A. Osako, Y. Pyrozhenko, P. Shermanau, and S. Federici, editors, *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Volume 5 Waste*, pages 6.1–6.72. IPCC, Switzerland, 2019. URL https://www.ipcc-nggip.iges.or.jp/public/2019rf/vol5.html. last access: 25 June 2023.

G. Van Drecht, A. F. Bouwman, J. Harrison, and J. M. Knoop. Global nitrogen and phosphate in urban wastewater for the period 1970 to 2050. *Global Biogeochemical Cycles*, 23(4):GB0A03, 2009. doi: 10.1029/2009GB003458.

UBA. *Einträge von Nähr- und Schadstoffen in die Oberflächengewässer.* German Environment Agency (UBA) [data set], 2020. URL https://www.umweltbundesamt.de/bild/stickstoff-phosphoreintraege-aus-punktquellen. last access: 26 January 2023.

P. Ebeling, R. Kumar, S. R. Lutz, T. Nguyen, F. Sarrazin, M. Weber, O. Büttner, S. Attinger, and A. Musolff. Quadica: water quality, discharge and catchment attributes for large-sample studies in germany. *Earth System Science Data*, 14(8):3715–3741, 2022. doi: 10.5194/essd-14-3715-2022.

M. Batool, F. J. Sarrazin, S. Attinger, N. B. Basu, K. Van Meter, and Kumar R. Long-term annual soil nitrogen surplus across Europe (1850–2019). *Scientific Data*, 9(612), 2022. doi: 10.1038/s41597-022-01693-9.

Review by Olga Vigiak

Overall comment: *Dear Authors, first of all many apologies for being late and delaying the review process. The manuscript describes an excellent dataset for nutrient emissions (N and P) to waters in Germany from 1950 to 2019. The authors compile a very thorough dataset that can be very useful. i have only minor comments for improving the manuscript.*

Reply: We thank the reviewer for their appreciation of our manuscript and for the useful comments and suggestions for improvement. In our replies below, line numbers refer to the revised version of our manuscript (version without track-changes).

Comment 1: *1) I like very much the uncertainty analysis, but I miss in the results a reporting of the final uncertainty of the parameters. I think this could be very interesting as it can affect similar analysis elsewhere. can you give revised intervals for the parameters and comment on the most sensitive ones? Referring only to a supplementary material part is too little in my opinion.*

Reply 1: Thank you for this suggestion. In our revised manuscript, we added an analysis regarding the reduction in the parameter ranges as suggested. We expanded on the analysis on the constraining of the parameter distributions that was presented in the first version of the manuscript and that revolves around Figure R1 reported below (which corresponds to Supplementary Fig. S35). Briefly, we found that the ranges could be substantially constrained in the behavioural parameter sample for four parameters. We added a table in the revised manuscript that reports the range reduction for these parameters (Table R1 reported below, which corresponds to Table 5 in the revised manuscript). Other parameters are also influential but their ranges cannot be reduced because of parameter interactions. We further expanded the paragraph that discusses the parameter uncertainty in Section 5.1 (L592–608) which now reads as follows:

"These differences in uncertainty bounds between the periods with and without observations can be interpreted in terms of parameter sensitivity. Table 5 shows that the parameter estimation procedure allows to appreciably reduce the ranges of a few parameters, namely (1) the efficiencies of tertiary treatment with targeted N and P removal ($eff_{3^N}^N$ and $eff_{3^P}^P$), and (2) the efficiencies of secondary and tertiary treatment without targeted N and P removal ($eff_{23^{noN}}^N$ and $eff_{23^{noP}}^P$) to a lesser extent and for some NUTS-1 regions only. This also reflects the fact that these two types of wastewater treatment tend to be more prevalent during the (recent) period with observations (Fig. 4). We observe that the posterior parameter ranges for $eff_{3^N}^N$ and $eff_{3^P}^P$ vary across NUTS-1 regions. Regarding $eff_{23^{noN}}^N$ and $eff_{23^{noP}}^P$, the higher values tend to be discarded in the posterior sample for some NUTS-1 regions. Notably, additional parameters have an impact on the simulated N and P outgoing load during the period with observations for at least some of the NUTS-1 regions. This regards mostly parameters that control the magnitude of the human N and P gross emissions (namely f_{waste}^{pro}, f_{pro}^N, and $f_{intake}^{N:P}$), the fractions of industrial/commercial to domestic N and P gross emissions in 2000 (namely $f_{ind:dom,2000}^N$ and

$f^{P}_{ind:dom,2000}$) and the fraction of N and P emissions lost during wastewater collection and transport ($f^{N,P}_{loss,transport}$). Although the ranges of these parameters cannot be reduced in the posterior parameter sample because of parameter interactions, their distributions are constrained (the posterior distributions deviate from the prior distributions in Supplementary Fig. S35). During the period without observations, parameters that could hardly be constrained may have a significant impact on the simulations. This for example can be the fractions of industrial/commercial to domestic N and P gross emissions in 1950 (namely $f^{N}_{ind:dom,1950}$ and $f^{P}_{ind:dom,1950}$) and the N and P removal efficiencies for primary treatment (eff^{N}_{1} and eff^{P}_{1}). Primary treatment was indeed widespread before 1990 (Fig. 4)."

Table R1: Lower and upper values of the parameters in the prior sample and in the posterior sample for each NUTS-1 region, and percentage reduction in the parameter ranges (Δ) in the posterior compared to the prior distribution (we report only the four parameters for which Δ is higher than 20% for at least one NUTS-1 region).

Parameter sample (prior or posterior for each NUTS-1 region)	eff^{N}_{23noN}			eff^{N}_{3N}			eff^{P}_{23noP}			eff^{P}_{3P}		
	Lower (-)	Upper (-)	Δ (%)	Lower (-)	Upper (-)	Δ (%)	Lower (-)	Upper (-)	Δ (%)	Lower (-)	Upper (-)	Δ (%)
Prior distribution	0.35	0.60	-	0.70	0.95	-	0.45	0.65	-	0.80	0.98	-
Schleswig-Holstein	0.35	0.51	36.4	0.87	0.95	68.2	0.45	0.56	45.8	0.88	0.98	44.7
Hamburg	0.36	0.60	2.3	0.73	0.82	63.1	0.45	0.65	1.7	0.89	0.92	79.2
Lower Saxony	0.35	0.60	1.3	0.88	0.93	83.5	0.46	0.65	3.4	0.90	0.94	81.2
Bremen	0.35	0.60	1.2	0.76	0.84	68.4	0.45	0.65	1.4	0.92	0.95	86.8
North Rhine-Westphalia	0.35	0.43	69.2	0.80	0.88	69.0	0.45	0.65	1.0	0.88	0.93	67.6
Hesse	0.35	0.44	63.5	0.77	0.85	69.3	0.45	0.64	6.0	0.80	0.89	49.5
Rhineland-Palatinate	0.35	0.52	30.5	0.82	0.89	71.6	0.45	0.62	17.0	0.86	0.92	65.0
Baden-WÃ¼rttemberg	0.35	0.60	1.4	0.71	0.81	60.5	0.45	0.55	50.4	0.85	0.91	67.1
Bavaria	0.35	0.47	51.8	0.70	0.80	63.3	0.45	0.65	1.8	0.80	0.88	53.4
Saarland	0.35	0.60	1.7	0.76	0.86	59.3	0.45	0.65	2.2	0.80	0.83	80.9
Berlin	0.35	0.60	2.8	0.83	0.90	73.5	0.45	0.64	6.0	0.96	0.97	91.7
Brandenburg	0.36	0.60	3.5	0.85	0.91	76.6	0.48	0.64	17.7	0.94	0.96	86.4
Mecklenburg-Vorpommern	0.35	0.60	1.8	0.89	0.93	81.5	0.45	0.64	6.3	0.91	0.96	71.7
Saxony	0.36	0.60	3.0	0.75	0.84	65.3	0.45	0.65	4.0	0.83	0.89	65.2
Saxony-Anhalt	0.35	0.60	1.5	0.87	0.92	80.9	0.47	0.65	9.1	0.89	0.93	75.0
Thuringia	0.35	0.60	1.9	0.86	0.93	72.5	0.45	0.58	34.2	0.87	0.94	62.2

The parameters are defined in Table 1. The prior parameter sample is the same for each NUTS-1 region. For a given parameter, the percentage reduction in the range (Δ (%)) is calculated as $\Delta = 100(1 - \frac{Upper_{posterior} - Lower_{posterior}}{Upper_{prior} - Lower_{prior}})$, where $Upper_{posterior}$ and $Lower_{posterior}$ are the upper and lower values respectively in the posterior parameter sample, and $Upper_{priori}$ and $Lower_{prior}$ are the upper and lower values respectively in the prior parameter sample.

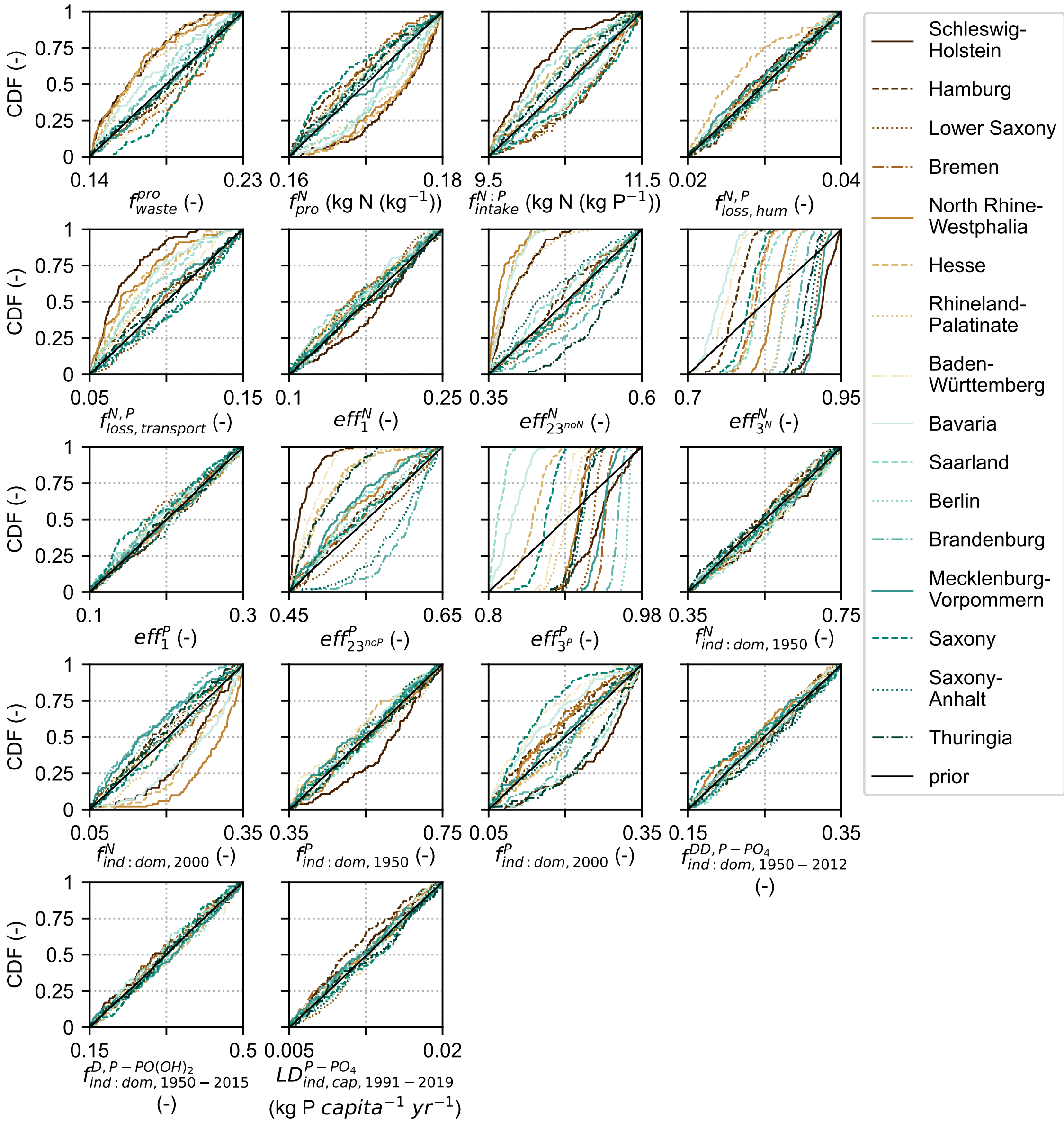

Figure R1: Cumulative distribution functions (CDFs) of the 18 point sources model parameters at NUTS-1 level. Colored lines refer to the posterior CDFs resulting from the 100 posterior parameterisations, and grey lines refer to the prior CDFs in the original sample of size 100,000. The prior CDFs are all uniform and are identical for all NUTS-1 regions. Parameters for which the posterior distributions is different from the prior distribution are influential for the metrics used for parameter estimation (here root mean square error), for instance the efficiency of tertiary treatment with N and P removal (eff_3^N and eff_3^P). However, the fact that the distributions are the same (for instance the industrial/commercial detergent parameters $f_{ind:dom,1950-2012}^{DD,P-PO_4}$, $f_{ind:dom,1950-2015}^{D,P-PO(OH)_2}$, and $LD_{ind,cap,1991-2019}^{P-PO_4}$) does not necessarily mean that the parameter is not influential, as it could have an impact through interactions with other parameters.

Comment 2: *2) the paper is very long, and at times verbose, and should be summarized more, particularly for the Material and methods part - we get to results only at page 27. I understand the need for details, and appreciate the work behind collecting all the data, but i think maybe section 3 on data can be shortened. Also, some sentences can be simplified: for example, authors could more simply refer to the many supplementary figures by mentioning them in brackets instead of saying 'as reported in...' etc.*

Reply 2: Thank you for this assessment. We agree that the manuscript should be more concise. We followed your suggestion and significantly reduced the Section 3 on data (by about two pages). We removed nonessential details from the main manuscript and moved them into the supplements. We improved the readability of the results section (including in response to Comment 7 below) and also referred to supplementary materials in a more concise manner throughout the manuscript.

Comment 3: *I have a question with regards to section 3.5.1 (maybe I missed it): the EU dataset reports for some UWWTPs incoming and outgoing N and P content, but usually this is only a smaller subsample of all UWWTPs reported in the database. are these the values that were used? what is the % of WWTPs that reported these data? Further, the EU dataset now reports also 2018 (and 2020), these data were not used (why)?*

Reply 3: Thank you for these questions. We could only create a dataset of WWTP load data for the year 2016 because we combined both the EU UWWTP dataset and a dataset specific to Germany (that we call "DE dataset") that only provides load data in 2015 or 2016 depending on the plants. It is important to consider the DE dataset, since it also includes smaller WWTPs that treat the wastewater in agglomerations smaller than 2000 Population Equivalent (PE).
Regarding the reporting of load data in the EU UWWTP dataset, it is indeed often a small subsample of all WWTPs reported in the database. However, for Germany the data are relatively more complete. We also combine the EU UWWTP dataset with the DE dataset that reports the load for some of the larger WWTPs that are present in the EU dataset. We demonstrate that the impact of the missing values on the total load at NUTS-1 level in our combined EU/DE dataset is limited, as detailed in the following.
We filled the missing values of the N and P load in our combined dataset using values for the other years provided in the EU dataset when available or we performed extrapolation based on the PE. Our final dataset comprises values of the incoming and outgoing N and P loads for over 99 % of the 9006 WWTPs, while only for 39 WWTPs no values at all of the N and P loads can be estimated and for 16 WWTPs no value of the incoming N and P loads and the outgoing N loads can be derived. This is shown in dark red color in Fig R2 reported below (which is Supplementary Figure S33). In this figure, we see that the number of WWTPs for which the values of the load have to be estimated based on PE can be large, in particular for the outgoing N load for the NUTS-1 region of Thuringia or more in general for the N and P incoming load for all NUTS-1 regions (light red color in Fig R2). However, the corresponding values of the loads (light red color in Fig R3, which is Supplementary Fig. S34) are actual small compared to the values that are derived from data (dark blue color in Fig R3). We find that these filled

values account for less than 8 % of the total load at NUTS-1 level, and have therefore an overall limited impact on our combined dataset.

In our revised mansucript, we substantially revised the data section, including section 3.5.1 to improve readability in response to Comment 2 above and to clarify the questions raised by the Reviewer in Comment 3. We refer Supplementary Sect. S9 for Details regarding the WWTP data .

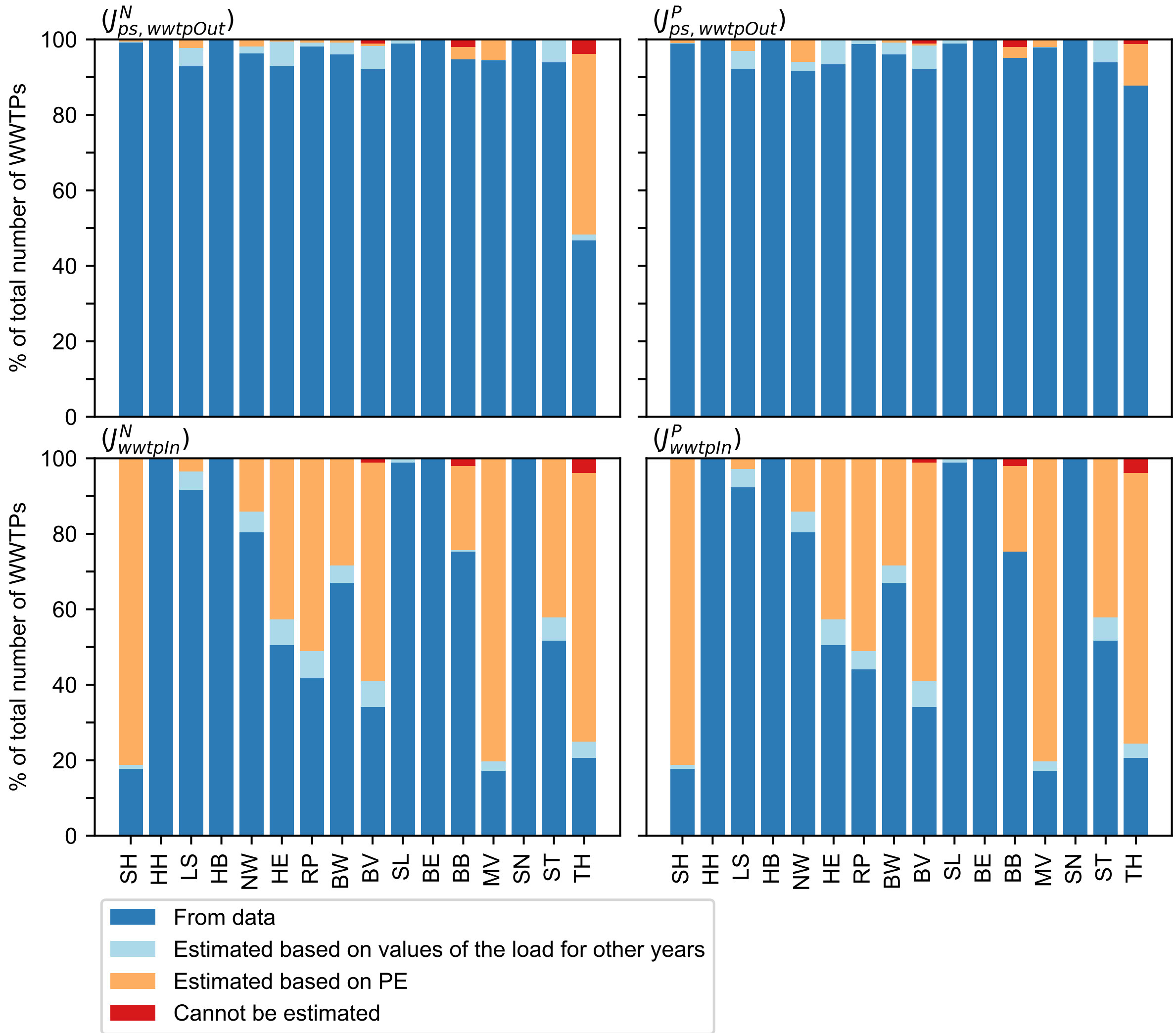

Figure R2: Distribution of the WWTPs in the combined EU/DE dataset according to the source of the load data at NUTS-1 level. Specifically, the figure shows the percentage of WWTPs in the combined dataset for which (1) the data provide the value of the load for the year 2016 (or 2015) either in the EU or in the DE dataset (dark blue), (2) the value of the load is filled based on values for the other years provided in the EU dataset (light blue), (3) the value of the load is filled based on the value of the entering load expressed in population equivalent (PE; light red), and (4) the value of the load cannot be estimated (dark red). The figure reports these percentages for the outgoing N load ($J^N_{ps,wwtpOut}$), the outgoing P load ($J^P_{ps,wwtpOut}$), the incoming N load (J^N_{wwtpIn}), and the incoming P load (J^P_{wwtpIn}), and for the 16 NUTS-1 regions of Germany.

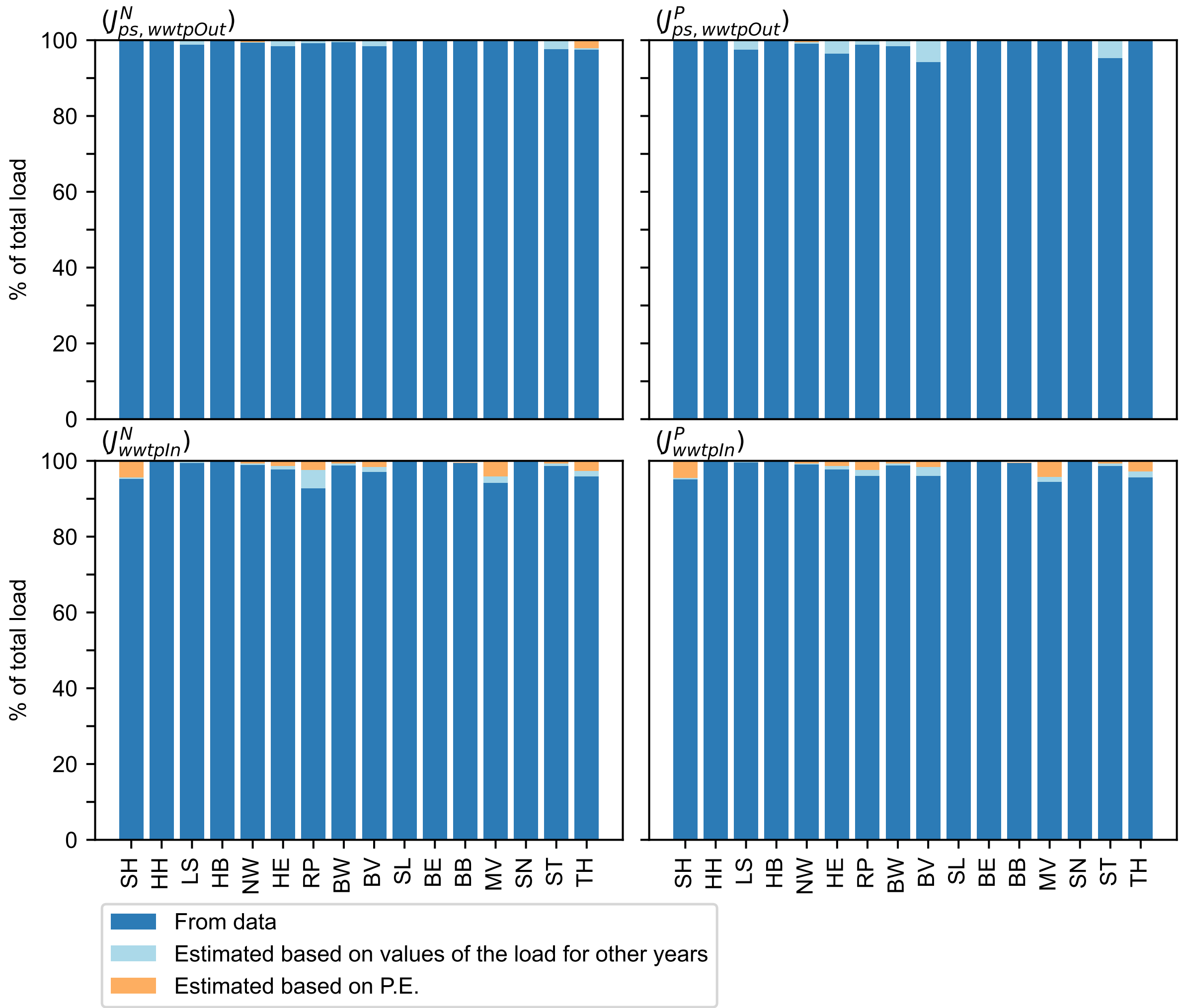

Figure R3: Distribution of the load in the combined EU/DE dataset according to the source of the load data at NUTS-1 level. Specifically, percentage of the load in the combined dataset that come from (1) data for the year 2016 (or 2015) provided either in the EU dataset or in the DE dataset (dark blue), (2) estimates based on values for the other years provided in the EU dataset (light blue), (3) estimates based on the value of the entering load expressed in population equivalent (PE; light red). The figure reports these percentages for the outgoing N load ($J^N_{ps,wwtpOut}$), the outgoing P load ($J^P_{ps,wwtpOut}$), the incoming N load (J^N_{wwtpIn}), and the incoming P load (J^P_{wwtpIn}), and for the 16 NUTS-1 regions of Germany. The figure was drawn using the average of the upper and lower bound load estimates. For each NUTS-1 region, the values of the load that come from data account for at least 97.5 %, 94.2 %, 92.8 %, 94.5 % of the total value of $J^N_{ps,wwtpOut}$, $J^P_{ps,wwtpOut}$, J^N_{wwtpIn}, and J^P_{wwtpIn}, respectively.

Comment 4: *other minor points: line87: i would remove 'Yes,'*

Reply 4: Thank you, we removed this typographical error.

Comment 5: *dashed lines in figure 2 and figure 7 are not dashed in printing, maybe use of a different color could help.*

Reply 5: Thank you for this comment. We changed the line style in Fig. 2 and Fig. 7. We wish to keep the color, since we used a colorblind safe scheme (`https://colorbrewer2.org/`).

Comment 6: *lines 403-404: the sentence starting with 'The assumption is supported.... ' is should be shortened and simplified*

Reply 6: Thank you. We have shortened and simplified this sentence which now reads as (L. 389–390): "The degree of urbanisation has a limited impact on the percentage of households and population owning a dishwasher (see Supplementary Sect. 5.2.3)."

Comment 7: *Paragraph starting at line 630 has some very long sentences that are difficult to read.*

Reply 7: Thank you for pointing this out. We realised that this paragraph is indeed too complicated and we simplified it in our revised manuscript (paragraph starting L. 575). In particular, we do not describe any more in details the results for the total point sources but only for the WWTP outgoing load. We refer to the supplements for details on the uncertainty bounds for the total point sources.

anyway, many congratulations for your excellent work!

Reply: Again, thank you very much for appreciation of our manuscript.